Снижение размерности пространства в задачах анализа временных рядов

Роман Исаченко

Научный руководитель:
д.ф.-м.н. В. В. Стрижов

Московский физико-технический институт
Факультет управления и прикладной математики
Кафедра «Интеллектуальные системы»

2020 г.

Задача декодирования временного ряда

Цель

Исследовать зависимости в пространствах объектов и ответов и построить устойчивую модель декодирования временных рядов в случае коррелированного описания данных.

Проблема

Целевая переменная – вектор, компоненты которого являются зависимыми.
Требуется построить модель, адекватно описывающую как пространство объектов так и пространство ответов при наблюдаемой мультикорреляции в обоих пространствах высокой размерности.

Решение

Для учёта зависимостей в пространствах объектов и ответов предлагается снизить размерность с использованием скрытого пространства.

Задача декодирования временного ряда

Цель

Исследуется фундаментальная проблема восстановления зависимостей не только в пространстве исходных описаний, но и в пространстве целевых переменных.

Задача

Предложить методы поиска скрытых закономерностей в связанных пространствах и алгоритмы снижения размерности, снижения сложности связанных моделей.

Решение

Линейные и нейросетевые модели, методы согласования связанных моделей в пространствах высокой размерности.

Картинка про работу

Литература

- Katrutsa A., Strijov V. Comprehensive study of feature selection methods to solve multicollinearity problem according to evaluation criteria // *Expert Systems with Applications* 76, 2017.
- Li J. et al. Feature selection: A data perspective //*ACM Computing Surveys (CSUR)* 50(6), 2017.
- Eliseyev A. et al. Iterative N-way partial least squares for a binary self-paced brain–computer interface in freely moving animals //*Journal of neural engineering* 4(8), 2011.
- Rodriguez-Lujan I. et al. Quadratic programming feature selection // *Journal of Machine Learning Research* 11(Apr), 2010.
- Motrenko A., Strijov V. Multi-way Feature Selection for ECoG-based Brain-Computer Interface // *Expert Systems with Applications* Submitted to the journal.

Постановка задачи

Модель

Назовём *моделью* параметрическую функцию $f : \mathbb{X} \times \mathbb{W} \to \mathbb{Y}$

- $\mathbb{X} \in \mathbb{R}^n$ – пространство независимой переменной;
- $\mathbb{Y} \in \mathbb{R}^r$ – пространство целевой переменной;
- $\mathbf{x} \in \mathbb{X}$ – реализация независимой переменной;
- $\mathbf{y} \in \mathbb{Y}$ – реализация целевой переменной;
- $\mathbb{W}$ – пространство параметров модели

Особенностью задачи является наличие избыточности описания как независимой переменной $\mathbf{x}$, так и целевой переменной $\mathbf{y}$.
То есть объекты $\mathbf{x}$ и $\mathbf{y}$ живут на некоторых многообразиях низкой размерности.
Примером таких многообразий могут являться линейными подпространствами.

Скрытое пространство

- Назовём пространство $\mathbb{T} \subset \mathbb{R}^l$ *скрытым пространством* для пространства $\mathbb{X} \in \mathbb{R}^n$ ($l \leqslant n$), если существуют *функция кодирования* $\varphi_e : \mathbb{X} \to \mathbb{T}$ и *функция декодирования* $\varphi_d : \mathbb{T} \to \mathbb{X}$ такие что

$$\forall \mathbf{x} \in \mathbb{X} \quad \exists \mathbf{t} \in \mathbb{T} : \varphi_d(\varphi_e(\mathbf{x})) = \varphi_d(\mathbf{t}) = \mathbf{x}.$$

- Аналогично введём определение *скрытого пространства* $\mathbb{U} \subset \mathbb{R}^s$ для целевого пространства $\mathbb{Y}$, *функции кодирования* $\psi_e : \mathbb{Y} \to \mathbb{U}$ и *декодирования* $\psi_d : \mathbb{U} \to \mathbb{Y}$

$$\forall \mathbf{y} \in \mathbb{Y} \quad \exists \mathbf{u} \in \mathbb{U} : \psi_d(\psi_e(\mathbf{y})) = \psi_d(\mathbf{u}) = \mathbf{y}.$$

- Назовём два скрытых пространства $\mathbb{T}$ и $\mathbb{U}$ *согласованными*, если существует *функция согласования* $h : \mathbb{T} \to \mathbb{U}$, такая что

$$\mathbf{y} = f(\mathbf{x}) = \psi_d(h(\varphi_e(x))).$$

Схема задачи декодирования

$$
\begin{tikzcd}
\mathbb{X} \subset \mathbb{R}^n \arrow[r, "f"] \arrow[d, bend right, "\varphi_e"'] & \mathbb{Y} \subset \mathbb{R}^r \arrow[d, bend left, "\psi_e"] \\
\mathbb{T} \subset \mathbb{R}^l \arrow[r, "h"] \arrow[u, bend right, "\varphi_d"'] & \mathbb{U} \subset \mathbb{R}^s \arrow[u, bend left, "\psi_d"]
\end{tikzcd}
$$

Многомерная регрессия

Дано

$(\mathbf{X}, \mathbf{Y})$ – выборка, $\mathbf{X} \in \mathbb{R}^{m \times n}$ – матрица объектов, $\mathbf{Y} \in \mathbb{R}^{m \times r}$ – матрица ответов,

$$\mathbf{X} = [\boldsymbol{\chi}_1, \dots, \boldsymbol{\chi}_n]; \quad \mathbf{Y} = [\boldsymbol{\nu}_1, \dots, \boldsymbol{\nu}_r].$$

Модель

$$\mathbf{y} = \boldsymbol{\Theta}\mathbf{x} + \varepsilon, \quad \boldsymbol{\Theta} \in \mathbb{R}^{r \times n}.$$

Функция потерь

$$\mathcal{L}(\boldsymbol{\Theta}|\mathbf{X}, \mathbf{Y}) = \left\| \underset{m \times r}{\mathbf{Y}} - \underset{m \times n}{\mathbf{X}} \cdot \underset{r \times n}{\boldsymbol{\Theta}}^{\mathsf{T}} \right\|_2^2 \to \min_{\boldsymbol{\Theta}}.$$

$$\boldsymbol{\Theta}^{\mathsf{T}} = (\mathbf{X}^{\mathsf{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathsf{T}}\mathbf{Y}.$$

Линейная зависимость столбцов матрицы $\mathbf{X}$ приводит к неустойчивому решению.

Для устранения сильной линейной зависимости предлагается использовать методы выбора признаков и снижения размерности пространства.

Снижение размерности пространства

Цель

- спроецировать исходные матрицы $\mathbf{X}$ и $\mathbf{Y}$ в общее латентное пространство;
- максимизировать ковариацию между образами;
- сохранить информацию об исходных матрицах.

Метод частных наименьших квадратов (PLS)

$$\underset{m\times n}{\mathbf{X}} = \underset{m\times l}{\mathbf{T}} \cdot \underset{l\times n}{\mathbf{P}^{\mathsf{T}}} + \underset{m\times n}{\mathbf{F}} = \sum_{k=1}^{l} \underset{m\times 1}{\mathbf{t}_k} \cdot \underset{1\times n}{\mathbf{p}_k^{\mathsf{T}}} + \underset{m\times n}{\mathbf{F}},$$

$$\underset{m\times r}{\mathbf{Y}} = \underset{m\times l}{\mathbf{U}} \cdot \underset{l\times r}{\mathbf{Q}^{\mathsf{T}}} + \underset{m\times r}{\mathbf{E}} = \sum_{k=1}^{l} \underset{m\times 1}{\mathbf{u}_k} \cdot \underset{1\times r}{\mathbf{q}_k^{\mathsf{T}}} + \underset{m\times r}{\mathbf{E}}.$$

$$\mathbf{U} \approx \mathbf{T}\mathbf{B}, \quad \mathbf{B} = \mathrm{diag}(\beta_k), \quad \beta_k = \mathbf{u}_k^{\mathsf{T}}\mathbf{t}_k / (\mathbf{t}_k^{\mathsf{T}}\mathbf{t}_k).$$

Псевдокод метода частных наименьших квадратов (PLS)

Require: $\mathbf{X}, \mathbf{Y}, l$;
Ensure: $\mathbf{T}, \mathbf{P}, \mathbf{Q}$;

normalize matrices $\mathbf{X}$ и $\mathbf{Y}$ by columns
initialize $\mathbf{u}_0$ (the first column of $\mathbf{Y}$)
$\mathbf{X}_1 = \mathbf{X}; \mathbf{Y}_1 = \mathbf{Y}$
for $k = 1, \ldots, l$ **do**
  repeat
    $\mathbf{w}_k := \mathbf{X}_k^\mathsf{T}\mathbf{u}_{k-1}/(\mathbf{u}_{k-1}^\mathsf{T}\mathbf{u}_{k-1}); \quad \mathbf{w}_k := \frac{\mathbf{w}_k}{\|\mathbf{w}_k\|}$
    $\mathbf{t}_k := \mathbf{X}_k\mathbf{w}_k$
    $\mathbf{c}_k := \mathbf{Y}_k^\mathsf{T}\mathbf{t}_k/(\mathbf{t}_k^\mathsf{T}\mathbf{t}_k); \quad \mathbf{c}_k := \frac{\mathbf{c}_k}{\|\mathbf{c}_k\|}$
    $\mathbf{u}_k := \mathbf{Y}_k\mathbf{c}_k$
  until $\mathbf{t}_k$ stabilizes
  $\mathbf{p}_k := \mathbf{X}_k^\mathsf{T}\mathbf{t}_k/(\mathbf{t}_k^\mathsf{T}\mathbf{t}_k), \ \mathbf{q}_k := \mathbf{Y}_k^\mathsf{T}\mathbf{t}_k/(\mathbf{t}_k^\mathsf{T}\mathbf{t}_k)$
  $\mathbf{X}_{k+1} := \mathbf{X}_k - \mathbf{t}_k\mathbf{p}_k^\mathsf{T}$
  $\mathbf{Y}_{k+1} := \mathbf{Y}_k - \mathbf{t}_k\mathbf{q}_k^\mathsf{T}$

Метод частных наименьших квадратов (PLS)

Утверждение (Исаченко, 2017)

Максимизация ковариации между векторами $\mathbf{t}_k$ и $\mathbf{u}_k$ приводит к наилучшему описанию матриц $\mathbf{X}$ и $\mathbf{Y}$ с учётом их взаимосвязи.

Утверждение (Исаченко, 2017)

Вектора $\mathbf{w}_k$ и $\mathbf{c}_k$ – собственные вектора матриц $\mathbf{X}_k^{\mathsf{T}}\mathbf{Y}_k\mathbf{Y}_k^{\mathsf{T}}\mathbf{X}_k$ и $\mathbf{Y}_k^{\mathsf{T}}\mathbf{X}_k\mathbf{X}_k^{\mathsf{T}}\mathbf{Y}_k$, соответствуюшие максимальным собственным значениям.

Утверждение (Исаченко, 2017)

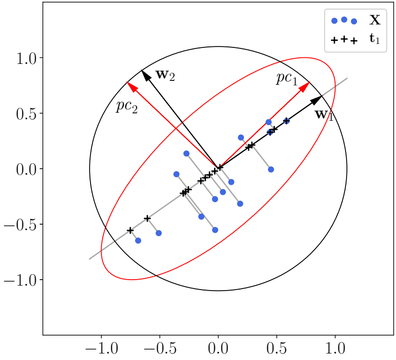
Правила обновления векторов (6)–(9) максимизируют ковариацию между векторами $\mathbf{t}_k$ и $\mathbf{u}_k$.

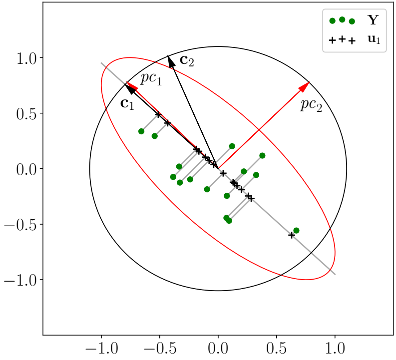
Модель PLS регрессии

$$\mathbf{Y} = \mathbf{U}\mathbf{Q}^{\mathsf{T}} + \mathbf{E} \approx \mathbf{T}\mathbf{B}\mathbf{Q}^{\mathsf{T}} + \mathbf{E} = \mathbf{X}\mathbf{W}^*\mathbf{B}\mathbf{Q}^{\mathsf{T}} + \mathbf{E} = \mathbf{X}\boldsymbol{\Theta} + \mathbf{E}.$$

$$\boldsymbol{\Theta} = \mathbf{W}(\mathbf{P}^{\mathsf{T}}\mathbf{W})^{-1}\mathbf{B}\mathbf{Q}^{\mathsf{T}}, \quad \mathbf{T} = \mathbf{X}\mathbf{W}^*, \quad \text{where } \mathbf{W}^* = \mathbf{W}(\mathbf{P}^{\mathsf{T}}\mathbf{W})^{-1}.$$

Пример PLS регрессии в двумерном случае





Задача выбора признаков

Требуется

Найти бинарный вектор $\mathbf{a} = \{0, 1\}^n$, компоненты – индикаторы выбранных признаков.

Функция ошибки отбора признаков

$$\mathbf{a} = \underset{\mathbf{a}' \in \{0,1\}^n}{\arg\min}\, S(\mathbf{a}' | \mathbf{X}, \mathbf{Y}).$$

Релаксация

Замена дикретной области определения $\{0, 1\}^n$ на непрерывную релаксацию $[0, 1]^n$:

$$\mathbf{z} = \underset{\mathbf{z}' \in [0,1]^n}{\arg\min}\, S(\mathbf{z}' | \mathbf{X}, \mathbf{Y}), \quad a_j = [z_j > \tau].$$

Получив **a**, решаем задачу регрессии:

$$\mathcal{L}(\boldsymbol{\Theta}_\mathbf{a} | \mathbf{X}_\mathbf{a}, \mathbf{Y}) = \left\| \mathbf{Y} - \mathbf{X}_\mathbf{a} \boldsymbol{\Theta}_\mathbf{a}^\mathsf{T} \right\|_2^2 \to \min_{\boldsymbol{\Theta}_\mathbf{a}},$$

где индекс **a** обозначает подматрицу с номерами столбцов, для которых $a_j = 1$.

Quadratic Programming Feature Selection

$$\|\boldsymbol{\nu} - \mathbf{X}\boldsymbol{\theta}\|_2^2 \to \min_{\boldsymbol{\theta} \in \mathbb{R}^n}.$$

Задача квадратичного программирования

$$S(\mathbf{z}|\mathbf{X}, \boldsymbol{\nu}) = (1 - \alpha) \cdot \underbrace{\mathbf{z}^{\mathsf{T}} \mathbf{Q} \mathbf{z}}_{\mathrm{Sim}(\mathbf{X})} - \alpha \cdot \underbrace{\mathbf{b}^{\mathsf{T}} \mathbf{z}}_{\mathrm{Rel}(\mathbf{X}, \boldsymbol{\nu})} \to \min_{\substack{\mathbf{z} \geqslant \mathbf{0}_n \\ \mathbf{1}_n^{\mathsf{T}} \mathbf{z} = 1}}.$$

- $\mathbf{z} \in [0, 1]^n$ – значимость признаков;
- $\mathbf{Q} \in \mathbb{R}^{n \times n}$ – матрица парных взаимодействий признаков;
- $\mathbf{b} \in \mathbb{R}^n$ – вектор релевантностей признаков к целевой переменной.

$$\mathbf{Q} = \left[\left| \mathrm{corr}(\boldsymbol{\chi}_i, \boldsymbol{\chi}_j) \right| \right]_{i,j=1}^n, \quad \mathbf{b} = \left[\left| \mathrm{corr}(\boldsymbol{\chi}_i, \boldsymbol{\nu}) \right| \right]_{i=1}^n.$$

Утверждение (Исаченко, 2018)

В случае полуопределенной матрицы $\mathbf{Q}$ *задача QPFS является выпуклой. Полуопределенная релаксация – сдвиг спектра:*

$$\mathbf{Q} \to \mathbf{Q} - \lambda_{\min} \mathbf{I}.$$

Многомерный QPFS

Агрегирование релевантностей по целевым векторам (RelAgg)

$$\mathbf{b} = [|\text{corr}(\boldsymbol{\chi}_i, \boldsymbol{\nu})|]_{i=1}^n \to \mathbf{b} = \left[\sum_{k=1}^{r} |\text{corr}(\boldsymbol{\chi}_i, \boldsymbol{\nu}_k)|\right]_{i=1}^n.$$

Недостаток: нет учёта зависимостей в матрице $\mathbf{Y}$.

Симметричный учёт значимостей (SymImp)

Штрафуем коррелированные целевые вектора с помощью $\text{Sim}(\mathbf{Y})$

$$\alpha_1 \cdot \underbrace{\mathbf{z}_x^{\mathsf{T}} \mathbf{Q}_x \mathbf{z}_x}_{\text{Sim}(\mathbf{X})} - \alpha_2 \cdot \underbrace{\mathbf{z}_x^{\mathsf{T}} \mathbf{B} \mathbf{z}_y}_{\text{Rel}(\mathbf{X}, \mathbf{Y})} + \alpha_3 \cdot \underbrace{\mathbf{z}_y^{\mathsf{T}} \mathbf{Q}_y \mathbf{z}_y}_{\text{Sim}(\mathbf{Y})} \to \min_{\substack{\mathbf{z}_x \geqslant \mathbf{0}_n,\, \mathbf{1}_n^{\mathsf{T}} \mathbf{z}_x = 1 \\ \mathbf{z}_y \geqslant \mathbf{0}_r,\, \mathbf{1}_r^{\mathsf{T}} \mathbf{z}_y = 1}}.$$

$$\mathbf{Q}_x = \left[|\text{corr}(\boldsymbol{\chi}_i, \boldsymbol{\chi}_j)|\right]_{i,j=1}^n, \quad \mathbf{Q}_y = [|\text{corr}(\boldsymbol{\nu}_i, \boldsymbol{\nu}_j)|]_{i,j=1}^r, \quad \mathbf{B} = [|\text{corr}(\boldsymbol{\chi}_i, \boldsymbol{\nu}_j)|]_{\substack{i=1,\dots,n \\ j=1,\dots,r}}.$$

$$\alpha_1 + \alpha_2 + \alpha_3 = 1, \quad \alpha_i \geqslant 0.$$

Многомерный QPFS

SymImp штрафует коррелированные целевые вектора, которые в меньшей мере объясняются признаками.

$$\alpha_1 \cdot \underbrace{\mathbf{z}_x^{\mathsf{T}} \mathbf{Q}_x \mathbf{z}_x}_{\mathrm{Sim}(\mathbf{X})} - \alpha_2 \cdot \underbrace{\mathbf{z}_x^{\mathsf{T}} \mathbf{B} \mathbf{z}_y}_{\mathrm{Rel}(\mathbf{X},\mathbf{Y})} \to \min_{\substack{\mathbf{z}_x \geqslant \mathbf{0}_n \\ \mathbf{1}_n^{\mathsf{T}} \mathbf{z}_x = 1}}; \quad \alpha_3 \cdot \underbrace{\mathbf{z}_y^{\mathsf{T}} \mathbf{Q}_y \mathbf{z}_y}_{\mathrm{Sim}(\mathbf{Y})} + \alpha_2 \cdot \underbrace{\mathbf{z}_x^{\mathsf{T}} \mathbf{B} \mathbf{z}_y}_{\mathrm{Rel}(\mathbf{X},\mathbf{Y})} \to \min_{\substack{\mathbf{z}_y \geqslant \mathbf{0}_r \\ \mathbf{1}_r^{\mathsf{T}} \mathbf{z}_y = 1}}.$$

Минимаксный подход (MinMax / MaxMin)

$$\min_{\substack{\mathbf{z}_x \geqslant \mathbf{0}_n \\ \mathbf{1}_n^{\mathsf{T}} \mathbf{z}_x = 1}} \max_{\substack{\mathbf{z}_y \geqslant \mathbf{0}_r \\ \mathbf{1}_r^{\mathsf{T}} \mathbf{z}_y = 1}} \left(\text{or} \max_{\substack{\mathbf{z}_y \geqslant \mathbf{0}_r \\ \mathbf{1}_r^{\mathsf{T}} \mathbf{z}_y = 1}} \min_{\substack{\mathbf{z}_x \geqslant \mathbf{0}_n \\ \mathbf{1}_n^{\mathsf{T}} \mathbf{z}_x = 1}} \right) \left[\alpha_1 \cdot \underbrace{\mathbf{z}_x^{\mathsf{T}} \mathbf{Q}_x \mathbf{z}_x}_{\mathrm{Sim}(\mathbf{X})} - \alpha_2 \cdot \underbrace{\mathbf{z}_x^{\mathsf{T}} \mathbf{B} \mathbf{z}_y}_{\mathrm{Rel}(\mathbf{X},\mathbf{Y})} - \alpha_3 \cdot \underbrace{\mathbf{z}_y^{\mathsf{T}} \mathbf{Q}_y \mathbf{z}_y}_{\mathrm{Sim}(\mathbf{Y})} \right].$$

Теорема (Исаченко, 2018)

Для положительно определенных матриц $\mathbf{Q}_x$ и $\mathbf{Q}_y$ minmax и maxmin задачи достигают одинакового значения функционала.

Теорема (Исаченко, 2018)

Минимаксная задача эквивалентна задаче квадратичного программирования с $n + r + 1$ переменными.

Для получения выпуклой задачи применяется сдвиг спектра.

Многомерный QPFS

Максимизация релевантностей (MaxRel)

$$\min_{\substack{\mathbf{z}_x \geqslant \mathbf{0}_n \\ \mathbf{1}_n^{\mathsf{T}} \mathbf{z}_x = 1}} \max_{\substack{\mathbf{z}_y \geqslant \mathbf{0}_r \\ \mathbf{1}_r^{\mathsf{T}} \mathbf{z}_y = 1}} \left[(1-\alpha) \cdot \mathbf{z}_x^{\mathsf{T}} \mathbf{Q}_x \mathbf{z}_x - \alpha \cdot \mathbf{z}_x^{\mathsf{T}} \mathbf{B} \mathbf{z}_y \right].$$

Теорема (Исаченко, 2018)

Для положительно определенной матрицы $\mathbf{Q}_x$ minmax и maxmin задачи достигают одинакового значения функционала.

Асимметричный учёт значимостей (AsymImp)

$$\alpha_1 \cdot \underbrace{\mathbf{z}_x^{\mathsf{T}} \mathbf{Q}_x \mathbf{z}_x}_{\mathrm{Sim}(\mathbf{X})} - \alpha_2 \cdot \underbrace{\left(\mathbf{z}_x^{\mathsf{T}} \mathbf{B} \mathbf{z}_y - \mathbf{b}^{\mathsf{T}} \mathbf{z}_y \right)}_{\mathrm{Rel}(\mathbf{X}, \mathbf{Y})} + \alpha_3 \cdot \underbrace{\mathbf{z}_y^{\mathsf{T}} \mathbf{Q}_y \mathbf{z}_y}_{\mathrm{Sim}(\mathbf{Y})} \to \min_{\substack{\mathbf{z}_x \geqslant \mathbf{0}_n, \, \mathbf{1}_n^{\mathsf{T}} \mathbf{z}_x = 1 \\ \mathbf{z}_y \geqslant \mathbf{0}_r, \, \mathbf{1}_r^{\mathsf{T}} \mathbf{z}_y = 1}}.$$

При $b_j = \max\limits_{i=1,\dots n} [\mathbf{B}]_{i,j}$ коэффициенты при $\mathbf{z}_y$ в $\mathrm{Rel}(\mathbf{X}, \mathbf{Y})$ неотрицательны.

Утверждение (Исаченко, 2017)

В одномерном случае $r = 1$ предлагаемые стратегии SymImp, MinMax, MaxMin, MaxRel, AsymImp совпадают с исходным алгоритмом QPFS.

Обобщение предложенных методов выбора признаков

Алгоритм	Критерий	Функция ошибки $S(\mathbf{z}\|\mathbf{X},\mathbf{Y})$
RelAgg	$\min\big[\mathrm{Sim}(\mathbf{X}) - \mathrm{Rel}(\mathbf{X},\mathbf{Y})\big]$	$\min\limits_{\mathbf{z}_x}\big[(1-\alpha)\cdot\mathbf{z}_x^{\mathsf{T}}\mathbf{Q}_x\mathbf{z}_x - \alpha\cdot\mathbf{z}_x^{\mathsf{T}}\mathbf{B}\mathbf{1}_r\big]$
SymImp	$\min\big[\mathrm{Sim}(\mathbf{X}) - \mathrm{Rel}(\mathbf{X},\mathbf{Y}) + \mathrm{Sim}(\mathbf{Y})\big]$	$\min\limits_{\mathbf{z}_x,\,\mathbf{z}_y}\Big[\alpha_1\cdot\mathbf{z}_x^{\mathsf{T}}\mathbf{Q}_x\mathbf{z}_x - \alpha_2\cdot\mathbf{z}_x^{\mathsf{T}}\mathbf{B}\mathbf{z}_y + \alpha_3\cdot\mathbf{z}_y^{\mathsf{T}}\mathbf{Q}_y\mathbf{z}_y\Big]$
MinMax	$\min\big[\mathrm{Sim}(\mathbf{X}) - \mathrm{Rel}(\mathbf{X},\mathbf{Y})\big]$ $\max\big[\mathrm{Rel}(\mathbf{X},\mathbf{Y}) + \mathrm{Sim}(\mathbf{Y})\big]$	$\min\limits_{\mathbf{z}_x}\max\limits_{\mathbf{z}_y}\big[\alpha_1\cdot\mathbf{z}_x^{\mathsf{T}}\mathbf{Q}_x\mathbf{z}_x - \alpha_2\cdot\mathbf{z}_x^{\mathsf{T}}\mathbf{B}\mathbf{z}_y - \alpha_3\cdot\mathbf{z}_y^{\mathsf{T}}\mathbf{Q}_y\mathbf{z}_y\big]$
MaxRel	$\min\big[\mathrm{Sim}(\mathbf{X}) - \mathrm{Rel}(\mathbf{X},\mathbf{Y})\big]$ $\max\big[\mathrm{Rel}(\mathbf{X},\mathbf{Y})\big]$	$\min\limits_{\mathbf{z}_x}\max\limits_{\mathbf{z}_y}\big[(1-\alpha)\cdot\mathbf{z}_x^{\mathsf{T}}\mathbf{Q}_x\mathbf{z}_x - \alpha\cdot\mathbf{z}_x^{\mathsf{T}}\mathbf{B}\mathbf{z}_y\big]$
AsymImp	$\min\big[\mathrm{Sim}(\mathbf{X}) - \mathrm{Rel}(\mathbf{X},\mathbf{Y})\big]$ $\max\big[\mathrm{Rel}(\mathbf{X},\mathbf{Y}) + \mathrm{Sim}(\mathbf{Y})\big]$	$\min\limits_{\mathbf{z}_x,\mathbf{z}_y}\big[\alpha_1\mathbf{z}_x^{\mathsf{T}}\mathbf{Q}_x\mathbf{z}_x - \alpha_2\left(\mathbf{z}_x^{\mathsf{T}}\mathbf{B}\mathbf{z}_y - \mathbf{b}^{\mathsf{T}}\mathbf{z}_y\right) + \alpha_3\mathbf{z}_y^{\mathsf{T}}\mathbf{Q}_y\mathbf{z}_y\big]$

Внешние критерии качества

Нормированное RMSE

Качество предсказания:

$$\mathrm{sRMSE}(\mathbf{Y}, \widehat{\mathbf{Y}}_\mathbf{a}) = \sqrt{\frac{\mathrm{MSE}(\mathbf{Y}, \widehat{\mathbf{Y}}_\mathbf{a})}{\mathrm{MSE}(\mathbf{Y}, \overline{\mathbf{Y}})}} = \frac{\|\mathbf{Y} - \widehat{\mathbf{Y}}_\mathbf{a}\|_2}{\|\mathbf{Y} - \overline{\mathbf{Y}}\|_2}, \quad \text{где} \quad \widehat{\mathbf{Y}}_\mathbf{a} = \mathbf{X}_\mathbf{a}\mathbf{\Theta}_\mathbf{a}^{\mathsf{T}}.$$

$\overline{\mathbf{Y}}$ — константный прогноз.

Мультикорреляция

Среднее значение коэффициента множественной корреляции:

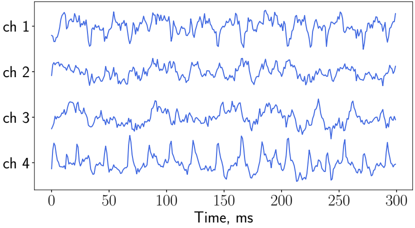
$$R^2 = \frac{1}{r}\mathrm{tr}\left(\mathbf{C}^{\mathsf{T}}\mathbf{R}^{-1}\mathbf{C}\right); \quad \mathbf{C} = [\mathrm{corr}(\boldsymbol{\chi}_i, \boldsymbol{\nu}_j)]_{\substack{i=1,\dots,n,\\ j=1,\dots,r}} \; \mathbf{R} = [\mathrm{corr}(\boldsymbol{\chi}_i, \boldsymbol{\chi}_j)]_{i,j=1}^{n}.$$

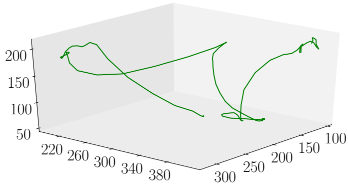
BIC

Компромисс между качеством предсказания и количеством выбранных признаков $\|\mathbf{a}\|_0$:

$$\mathrm{BIC} = m \ln\left(\mathrm{MSE}(\mathbf{Y}, \widehat{\mathbf{Y}}_\mathbf{a})\right) + \|\mathbf{a}\|_0 \cdot \log m.$$

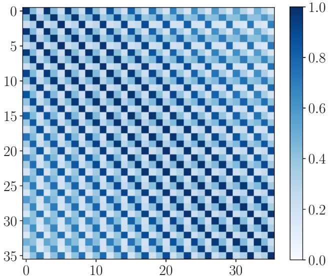
Вычислительный эксперимент, данные





$$\mathbf{X} \in \mathbb{R}^{m\times(32\cdot 27)}; \quad \mathbf{Y} \in \mathbb{R}^{m\times 3k}.$$

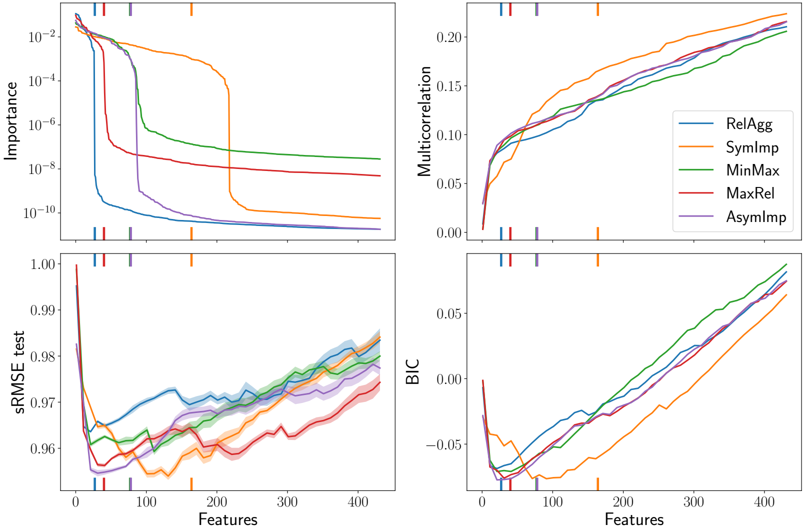
$$\mathbf{Y} = \begin{pmatrix} x_1\ y_1\ z_1 & \dots & x_k & y_k & z_k \\ x_2\ y_2\ z_2 & \dots & x_{k+1} & y_{k+1} & z_{k+1} \\ \dots & \dots & & \dots & \\ x_m\ y_m\ z_m & \dots & x_{m+k} & y_{m+k} & z_{m+k} \end{pmatrix}$$



Матрица корреляций $\mathbf{Y}$

http://neurotycho.org

Результаты эксперимента



Стабильность выбора признаков

Постановка эксперимента

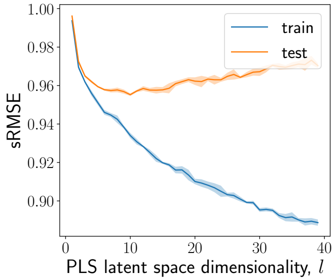
- создать бутстреп-выборки
$$(\mathbf{X}, \mathbf{Y}) \to \{(\mathbf{X}_1, \mathbf{Y}_1), \dots, (\mathbf{X}_s, \mathbf{Y}_s)\};$$
- решить задачу выбора признаков
$$\{(\mathbf{X}_1, \mathbf{Y}_1), \dots, (\mathbf{X}_s, \mathbf{Y}_s)\} \to \{\mathbf{z}_1, \dots, \mathbf{z}_s\};$$
- вычислить статистики
$$\{\mathbf{z}_1, \dots, \mathbf{z}_s\} \to \{\text{sRMSE}, \|\mathbf{a}\|_0, \text{Спирмен } \rho, \ell_2 \text{ расстояние}\}.$$

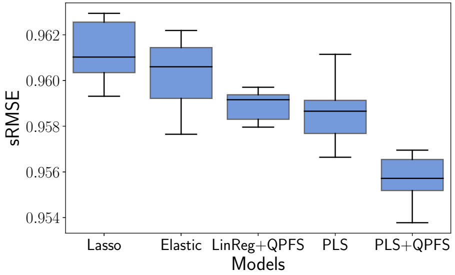
	sRMSE	$\|\mathbf{a}\|_0$	Спирмен ρ	ℓ_2 расстояние
RelAgg	0.965 ± 0.002	26.8 ± 3.8	0.915 ± 0.016	0.145 ± 0.018
SymImp	0.961 ± 0.001	224.4 ± 9.0	0.910 ± 0.017	0.025 ± 0.002
MinMax	0.961 ± 0.002	101.0 ± 2.1	0.932 ± 0.009	0.059 ± 0.004
MaxRel	0.958 ± 0.003	41.2 ± 5.2	0.862 ± 0.027	0.178 ± 0.010
AsymImp	0.955 ± 0.001	85.8 ± 10.2	0.926 ± 0.011	0.078 ± 0.007

QPFS vs PLS

Постановка эксперимента

Сравнить отбор признаков и снижение размерности пространства с помощью моделей линейной регрессии и PLS регрессии.





Результаты, выносимые на защиту

- Исследована задача декодирования сигналов в пространствах высокой размерности.
- Исследованы методы снижения размерности с анализом структуры пространства.
- Предложены методы для выбора признаков, учитывающие зависимости как в пространстве объектов, так и в пространстве ответов.
- Предложена комбинация методов выбора признаков и снижения размерности пространства.
- Создан макет системы, пригнозирующей сигналы в пространстве большой размерности.
- Предложенные алгоритмы выбора признаков доставляют устойчивые и адекватные решения в коррелированных пространствах высокой размерности.

Заключение

Публикации ВАК

- Исаченко Р.В., Стрижов В. В. Метрическое обучение в задачах мультиклассовой классификации временных рядов *Информатика и её применения*, 10(2), 2016.
- Isachenko R. et al. Feature Generation for Physical Activity Classification. *Artificial Intellegence and Decision Making*, 2018, подана в журнал.
- Isachenko R., Strijov V. Quadratic programming optimization for Newton method. *Lobachevskii Journal of Mathematics*, 2018, принята к публикации.
- Isachenko R., Vladimirova M., Strijov V. Dimensionality reduction for multivariate ECoG-based data. *Chemometrics*, 2018, готова к подаче.

Выступления с докладом

- Ломоносов, 2016, Москва. Метрическое обучение в задачах мультиклассовой классификации временных рядов.
- Intelligent Data Processing Conference, 2016, Барселона. Multimodel forecasting multiscale time series in internet of things.
- Математические методы распознавания образов ММРО, 2017, Таганрог. Локальные модели для классификации объектов сложной структуры.